# Bayes Theory: Risk and Reward

## The JCAT Computation Model

**John F. Lemmer, PhD**
**AFRL/IFSA**
john.lemmer@rl.af.mil

# Not all Bayes Tools are the Same

Theoretical soundness and accurate modeling ***you can use***!

# JCAT Goals

- Model Probabilistic Cause/Effect *over time*
- Maintain Semantic integrity
  - Probabilities IN
  - Probabilities OUT
  - Enable Model Analysis
- Use and leverage causal concepts e.g.
  - Synergy
  - Necessity ..
- Feasibility:
  - Model building
  - Computation

# JCAT Computational Model

- The primary contribution of the CAT research has been developing
  - A computational model for achieving CAT goals
  - Developing a user interface involving only SME type knowledge
- Utility of CAT goals is the 'Reward'
- Overcoming difficulties has been the risk.
  - The difficulties overcome are why not all "Bayes Tools" are created equally

# So Why Probabilities ?

- In a word: Semantics
  - Empirical Semantics
  - Rich Theory
- *Like the difference between qualitative and quantitative physics.*

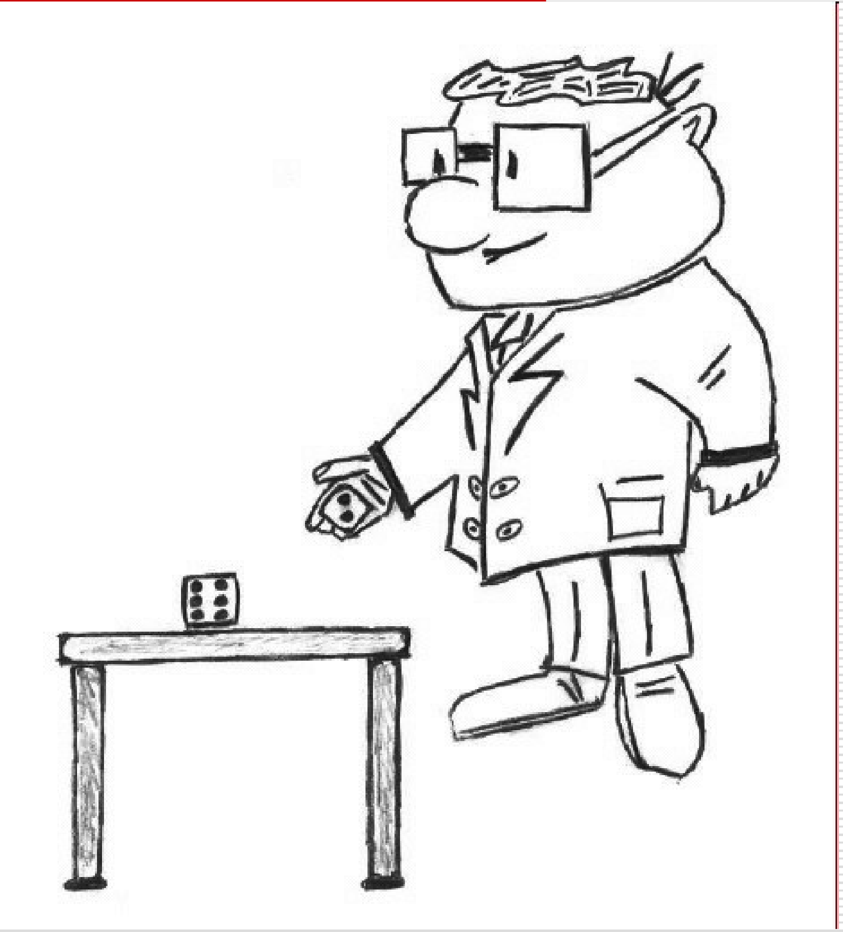
# The Rewards of Semantics

- Advantages of a theoretically sound foundation
  - Semantics
    - Inputs are well defined (unlike e.g. SIAM)
    - Outputs are well defined
  - Analysis
    - Vs. Prescription
    - Model acceptance/rejection

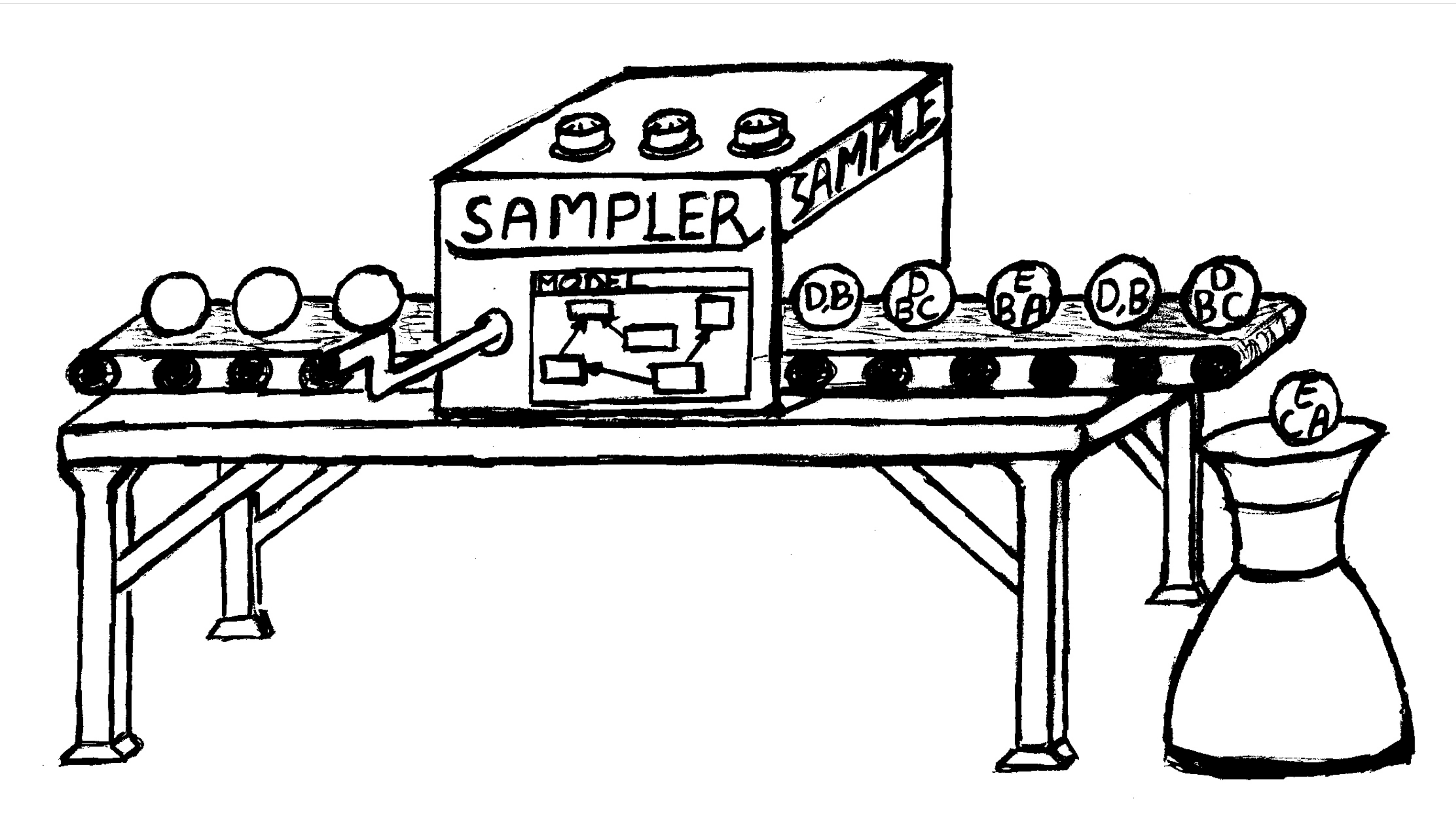
# Understanding the Risk

- What is Bayesian Analysis?
  - What is Bayesian probabilistic analysis?
  - What is Causal analysis?
  - Why are they hard?
- JCAT and its tradeoffs

# Dice



# JCAT Prediction



# Bayesian Inference

# Urn Model of Semantics

- Objective probabilities: Urn contains
  - Balls with labels e.g. any subset of {A,B,C,D,E}
  - Prediction is equivalent to rules for labeling the balls
  - Bayesian inference is equivalent to
    - Drawing one ball from an urn
    - Observing some of the labels; computing the probability of other labels on the same ball
  - Model verification
    - Likelihood that observed evidence is consistent with the model
- Subjective probabilities
  - Expert beliefs
  - Verified by model performance

# Being Bayesian is Hard

- Many ‘Bayesian’ tools are based on assumptions which
  - Destroy the semantics
    - e.g. After computation, parameters are not probabilities (except perhaps under extreme assumptions)
  - Limit model fidelity
  - Limit model analysis
- JCAT is based on more benign assumptions
  - As explained in the next few slides
    - Contrasted with alternate assumptions

# But what IS Bayesian Analysis?

# Textbook Bayes Rule

- Looks simple
- Very limited application
  - Only discrete events

# Textbook Bayes Rule

$$p'(q_i) = p(q_i/a) = \frac{p(q_i a)}{p(a)} = \frac{p(q_i/a)\,p(a)}{\sum_i p(a/q_i)\,p(q_i)}$$

If

$$p(a) = \sum_i p(a/q_i)\,p(q_i)$$

then the *q* must be disjoint, limiting the distributions which ca
For example the distribution in the previous slide cannot be mod

# What is the General Form of Bayes Rule?

- Very large arrays of numbers
  - e.g. more than $2^{100}$ in demo
- Thousands of user provided parameters

# BI Defined by Example

Evidence: p(a) = 1; p(~a) = 0
BR: redistribute prob. to match Evidence

Prior Probabilities

| ***c*** | ***b*** | ***a*** | ***p(.)*** |
|---|---|---|---|
| 0 | 0 | 0 | .05 |
| 0 | 0 | 1 | .12 |
| 0 | 1 | 0 | .1 |
| 0 | 1 | 1 | .15 |
| 1 | 0 | 0 | .08 |
| 1 | 0 | 1 | .2 |
| 1 | 1 | 0 | .2 |
| 1 | 1 | 1 | .1 |

Posterior Probabilities

| ***c*** | ***b*** | ***a*** | ***p(.)*** |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | .21 |
| 0 | 1 | 0 | 0 |
| 0 | 1 | 1 | .26 |
| 1 | 0 | 0 | 0 |
| 1 | 0 | 1 | .35 |
| 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | .18 |

# Early Developments

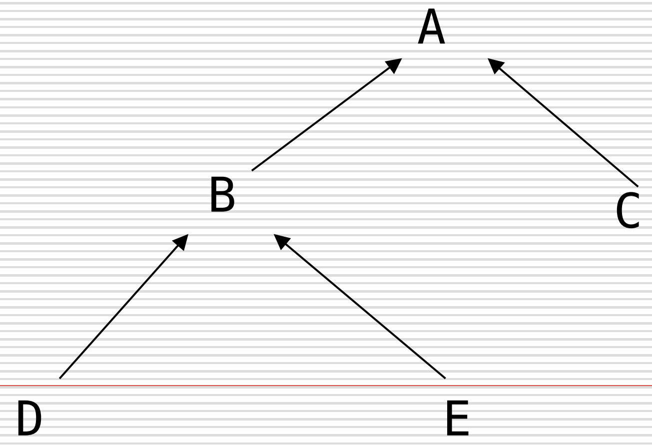
- Analogy developed between a Causal Model and a type of Markov model (subsequently know as a Bayes Net).
  - If the connections are sufficiently sparse, the so called “Junction Tree” algorithms give real traction on the computability problem
  - BTW modelling time usually destroys the sparseness.

# A (Markov Model) Bayesian Network

- {A,B,C,D,E} is the set of variables

$$p(A,B,C,D,E) = p(A/B,C,D,E)p(B/C,D,E)p(C/D,E)p($$

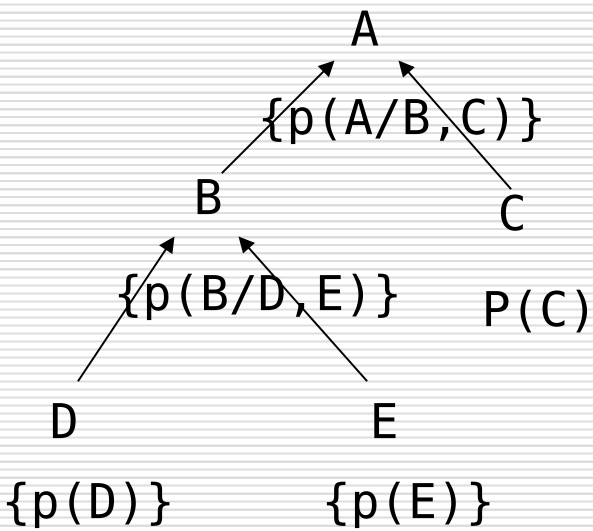
$$p(A,B,C,D,E) = p(A/B,C)p(B/D,E)p(C)p(D)p(E$$



# Bayes Nets

- Markov Model provides a simplified representation of the underlying distribution
- Markov Model
  - Can be justified by causal arguments
  - Conditional Probability Tables are sufficient for specification

# A Bayesian Network w/Conditional Probability Tables




- {A,B,C,D,E} is the set of variables

p(A,B,C,D,E) = p(A/B,C)p(B/D,E)p(C)p

# Modeler Tasks

- Build the graph model of causality
- Build Conditional Probability Tables
  - Full Specification (HUGIN, GENIE)
  - First Order
    - Causal Independence (e.g. SIAM)
    - Disjoint Causes (text book Bayes)
  - CAT approach: a compromise between causal independence and full specification
    - Specify ‘alone’ causation probability
    - Specify important groups of probabilities which are not causally independent
    - Algorithm estimates remaining groups to fill out the entire CPT

# Model Analysis

- Prediction
- Inference from Evidence
  - Given current evidence, predict nuclear capability as…
- Explanation
  - What is causing difference between now and then?
- Model acceptance/rejection
  - Less than 5% chance that evidence was drawn for a ‘substantially’ different model

# The End

# So Why Bayesian Probability?

- In a word: Semantics
  - Empirical Semantics
  - Empirical because:
    - Inputs *can be measured*
    - *Outputs can be measured*
    - *Computations result in semantic preserving, scientific predictions.*
- *Like the difference between qualitative and quantitative physics.*

# Rewards

- High quality models can be
  - Built feasibly
  - Results can be understood
  - Models can be analyzed

# Retrospective on AUAI

- In 1985, a workshop similar to this was held
  - Major issues included “Certainty Factors” (now long dead!) etc.
  - Resulted in an on-going professional association
- Since then, probabilities have taken over main stream AI e.g.
  - Text understanding
  - DARPA Grand Challenge
    - (see current Scientific American)